DOCUMENT RESUME

ED 161 900 TM 007 515

AUTHOR Sanner, Richard L.
TITLE Chi Square or K-S?--That is the Question.
PUB DATE [74]
NOTE 12p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Comparative Statistics; *Data Analysis; Goodness of Fit; *Rating Scales; Sampling; *Statistical Analysis; *Tests of Significance
IDENTIFIERS *Chi Square; *Kolmogorov Smirnov Test

ABSTRACT

Two types of statistical analyses of rating scale data are discussed. An example, with its accompanying mathematical calculations for each type, is presented; and the advantages and disadvantages of each method are compared. It is judged that the little-known and seldom-used Kolmogorov-Smirnov one-sample test should be reexamined because of its ease of use, retention of information, considerable power, and unique appropriateness for use with small samples. When the educatioral practitioner needs to analyze descriptive data, the Kolmogorov-Smirnov test should be given serious consideration. (Author/CTM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED161900

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# CHI SQUARE OR K-S? - THAT IS THE QUESTION

Richard L. Sanner, Ed.D.
Media Specialist
Associate Professor
Education Resource Center
Library
San Diego State University
San Diego, California 92182

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard Sanner

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM007 515

# CHI SQUARE OR K-S? - THAT IS THE QUESTION

RICHARD L. SANNER, ED.D.

Frequently educational practitioners find it necessary to engage in descriptive research involving rating scales such as opinionnaires and questionnaires. After such scales have been developed, validated, distributed, completed, and returned, the data collected must be subjected to statistical treatment to determine whether or not the findings are significant. In this article, two different types of analyses will be presented, their mathematical procedures will be compared, and the strengths and weaknesses of each will be discussed.

Before the researcher develops the rating instrument, a "straw man" type of hypothesis, the null hypothesis ($H_o$), must be formalized. The null hypothesis merely states that there is no difference between the numbers, or frequencies ($f$), of re-

sponses for each item of the instrument and a theoretical even distribution of such responses. The analysese of the data are then carried out to determine whether or not the null hypothesis can be rejected at some predetermined probability value (p) such as 5% or 1%. Once the mathematical calculations have been made, tables are referred to to determine such probabilities.

The most commonly used statistical treatment of such data, and the one referred to most frequently in the literature, is the chi-square ($X^2$) one-sample test. For each item of the instrument, the $X^2$ value is calculated and compared to a table of critical values of chi square to determine the level of significance. The calculations for $X^2$ are rather straight forward. They involve determining the dif-

ferences between the theoretical and observed frequencies for each item in the instrument, squaring these differences, adding these calculated values for all categories of each item, and dividing the result obtained by the theoretical frequency for each item. (Siegel, 1956.)

The second type of statistical treatment is the Kolmogorov-Smirnov (K-S) one-sample test. For each item of the instrument, the largest absolute cumulative difference (D) between observed and theoretical frequencies is compared to the table of critical values of D in the Kolmorgorov-Smirnov one-sample test to determine the level of significance. The calculations for D are also straight forward. They involve determining the cumulative frequencies, dividing these values by the total number of responses, subtracting from these computed values the

theoretical cumulative quantities under the $H_o$, and identifying the largest absolute value, D. The theoretical cumulative quantities under the $H_o$ are determined by the number of categories in each item of the instrument. For example, if there are five categories, the quantities will be .2 (1/5), .4 (2/5), .6 (3/5), .8 (4/5), and 1.0 (5/5); if four categories, the quantities will be .25 (1/4), .50 (2/4), .75 (3/4), and 1.00 (4/4); etc. (Siegel, 1956.)

A hypothetical case will illustrate the mathematical computations for the $X^2$ and K-S tests. The number of categories is five (1 = Very poor, 2 = Poor, 3 = Uncertain, 4 = Good, and 5 = Very good), the total number of responses is 25, and the returned opinionnaires contain the following frequencies: $f_1 = 1$, $f_2 = 2$, $f_3 = 5$, $f_4 = 10$, and $f_5 = 7$.

The theoretical frequencies under $H_o$, with five categories and 25 responses, would be five, (25/5).

The computations are, for $X^2$:

$$X^2 = \frac{(5-1)^2+(5-2)^2+(5-5)^2+(5-10)^2+(5-7)^2}{5}$$

$$= \frac{(4)^2+(3)^2+0+(-5)^2+(-2)^2}{5}$$

$$= \frac{16+9+0+25+4}{5}$$

$$= \frac{54}{5} = 10.8$$

and for K-S:

(Cummulative f: 1, 3, 8, 18, 25)

$$D_1 = \frac{1}{25} - .2 = .04 - .2 = -.16$$

$$D_2 = \frac{3}{25} - .4 = .12 - .4 = -.28$$

$$D_3 = \frac{8}{25} - .6 = .32 - .6 = -.28$$

$$D_4 = \frac{18}{25} - .8 = .72 - .8 = -.08$$

In the above illustration, the value for $X^2$ is found to be 10.8, while the value for D is found to be .28, the largest absolute deviation. By consulting the corresponding tables, it is determined that the level of significance for $X^2$ is less than .05 (5%) and for K-S it is also less than .05.

In order to carry out the calculations above, the formulas for $X^2$ and $\underline{D}$ were reduced to simplified forms and the values for the hypothetical case were plugged in. These procedures can be further simplified by the use of computers. In all probability most practitioners would not have access to the simplest of table-top computers, such as the Olivetti Underwood Programma 101. They must rely on calculators that will add, multiply, subtract, and divide. It will be readily seen that the use of such calculators would become quite time consuming when calculating $X^2$. The calculations of $\underline{D}$, however, are quite simple and quickly accomplished. Because of this, the simplest of the new pocket-size, battery-operated hand calculators can be used. It will be noted that when computing $\underline{D}$, the last frequency is not calculated for

the value will always be 0 (1 - 1). A definite advantage of the K-S method over the $X^2$ method is the general ease of computation with readily available equipment.

In the hypothetical case, there were 25 usable responses to that particular item which contained five categories. If there had not been at least that number of responses, the $X^2$ test could not be used. The requirement of the $X^2$ test is that no cell shall have fewer than five theoretical or expected frequencies. The only way this test can be used for fewer total responses is to reduce the number of categories. This can be accomplished by combining two or more categories into one new one. In the example, categories 1 and 2 and half of 3 could be combined into one category, and categories 4 and 5 with half of 3 in the other, thus reducing the total number of catego-

ries from five to two. This type of adjusting further complicates the computations for $X^2$. However, this disadvantage of the test for use with small samples can be of advantage when large samples are encountered.

Siegel (1956) points out that the power of the $X^2$ test, the probability of the test to reject the $H_o$ when it actually is false, is not known and not possible to determine. This is another disadvantage of $X^2$. On the other hand, the power of the K-S test is definitely superior to $X^2$ because it applies to small values, and categories need not be combined and thus lose information. Thus, the K-S test probably is a more powerful test than the $X^2$ test in all cases.

An earlier study recently reported by the author (Sanner, 1971; Sanner, 1974) involved the use of the $X^2$ test to detemine the significance

of the findings. A recalculation at the .05 or lower level of significance using the K-S test reveals that 79% were identical to the $X^2$ findings, 10% were lower, while 11% were larger.

In this article, two types of statistical analyses of rating scale data have been discussed; an example, with its accompanying mathematical calculations for each type, was then presented; and the advantages and disadvantages of each method were compared.

It would seem that the little-known and seldom-used Kolmogorov-Smirnov one-sample test be reexamined because of its ease of use, retention of information, considerable power, and unique ability of use with small samples. When the educational practitioner needs to analyse descriptive data, the K-S test should be given serious consideration.

REFERENCES

Sanner, Richard L. "An Evaluation of the Educational Media Programs of the California State Colleges." Unpublished Doctor's dissertation, Arizona State University, 1971.

Sanner, Richard L. "Evaluation of Educational Media Programs--by Faculty," Audiovisual Instruction, September 1974. Pp. 7-9.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.